

PATENT

## IN THE UNITED STATES

## PATENT AND TRADEMARK OFFICE

Official

APPLICANTS: Ted E. Dunning and Bradley D. Kindig

APPLICATION NO.: 09/848,982

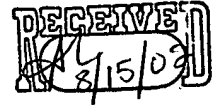
FILING DATE: May 3, 2001

TITLE: Text Equivalencing Engine

EXAMINER: Not yet known

GROUP ART UNIT: 2644

ATTY. DKT. NO.: 22227-05479



## CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Box Missing Parts, Commissioner For Patents, Washington, D.C. 20231, on the date shown below:

Dated: May 9, 2002

By:

Elaine M. Heal, Reg. No.: 44,149

COMMISSIONER FOR PATENTS  
WASHINGTON, D.C. 20231

## PETITION TO MAKE SPECIAL

SIR:

Applicants hereby petition to make the above-referenced application special as described in MPEP §708.02 VIII. No examination has yet been taken by an Examiner in this case. The claims are directed to a single invention. If the Office determines that all the claims presented are not obviously directed to a single invention, Applicants will make an election without traverse as a prerequisite to the grant of special status.

A pre-examination search has been made with respect to the present invention by the professional searching company MetroPatent. The following classes and subclasses were

**PATENT**

searched: Class 707, subclasses 532 and 540. Additional search of EAST was conducted by Applicants' representatives.

Pursuant to the provisions of 37 CFR 1.56 and 1.97-98, enclosed herewith is a modified form PTO-1449 listing references for consideration by the Examiner. A copy is enclosed of each of those references not already of record and deemed most closely related to the subject matter encompassed by the claims. The inclusion of these references shall not be construed as a representation regarding the completeness of the list of references, or that inclusion of a reference in this list is an admission that it is prior art or is pertinent to this application, or as an admission that the information listed is, or may be considered to be, material to patentability, or that no other material information exists, and shall not be construed as an admission against interest in any manner.

**Detailed Discussion of the References**

Pursuant to the provisions of MPEP §708.02 VIII, the following is a detailed discussion of the references listed in the accompanied modified form PTO-1449, which discussion points out, with the particularity required by 37 CFR 1.111(b) and (c), how the claimed subject matter is patentable over the references.

Independent Claim 1 recites:

1. A computer-implemented method of text equivalencing from a string of characters comprising:
  - modifying the string of characters using a predetermined set of heuristics;
  - comparing the modified string with a known string of characters in order to locate a match;

**PATENT**

responsive to not finding a match, forming a plurality of sub-strings of characters  
from the string of characters; and  
using an information retrieval technique on the sub-strings of characters to  
determine a known string of characters equivalent to the string of characters.

Independent claim 12 recites:

12. A computer implemented system for text equivalencing from a string of characters  
comprising:  
a heuristics module for modifying the string of characters using a predetermined set  
of heuristics;  
a comparator module, coupled to the heuristics module, for comparing the modified  
string with a known string of characters in order to find a match;  
a sub-string formation module, coupled to the comparator module, responsive to not  
finding a match, for forming a plurality of sub-strings of characters from the  
string of characters; and  
an information retrieval module, coupled to the sub-string formation module, for  
performing an information retrieval technique on the sub-strings of  
characters to determine a known string of characters equivalent to the string  
of characters.

Independent claim 23 recites:

**PATENT**

23. A computer-readable medium comprising computer-readable code for performing text equivalencing from a string of characters comprising:

- computer-readable code adapted to modify the string of characters using a predetermined set of heuristics;
- computer-readable code adapted to compare the modified string with a known string of characters in order to locate a match;
- computer-readable code, responsive to not finding a match, adapted to form a plurality sub-strings of characters from the string of characters; and
- computer-readable code adapted to use an information retrieval technique on the sub-strings of characters to determine a known string of characters equivalent to the string of characters.

Independent claim 34 recites:

34. A computer-implemented system for performing text equivalencing from a string of characters comprising:

- a modifying means for modifying the string of characters using a predetermined set of heuristics;
- a comparator means for comparing the modified string with a known string of characters in order to locate a match;
- responsive to not finding a match, a formation means for forming a plurality sub-strings of characters from the string of characters; and

**PATENT**

an information retrieval means for determining a known string of characters equivalent to the string of characters.

The present invention thus provides a text equivalencing engine capable of determining when one string of characters or text is equivalent or probably equivalent to another string of characters or text. In one embodiment, the string of characters is textual information describing a track of music such as a song title, artist name, album name or any combination of these attributes. In another embodiment, the string of characters is textual information describing a book or magazine such as title of the book or magazine, author name, or publisher. In another embodiment, the text is a genetic sequence or the text is a computer program listing. One skilled in the art will recognize that the techniques of the present invention may be applied to any type of text or character string. The present invention can determine when two strings of characters are equivalent even when there are misspellings or typographical errors that are not addressed by a set of heuristics. The present invention can be used to accurately perform text equivalencing in most cases. Thus, the present invention overcomes the above-described problems of misspellings and typographical errors.

The claims are directed toward several different variations of these text equivalencing techniques. Claim 1 recites a method for text equivalencing by modifying a string of characters, comparing the modified string with a known string in order to locate a match, when a match is not found, forming sub-strings of characters, and using an information retrieval technique on the sub-strings to determine a known string of characters equivalent to the string. Claim 12 recites a system claim corresponding to the method of claim 1. Claim 23 recites a computer program product corresponding to the method of claim 1. Claim 34 recites

**PATENT**

a system for text equivalencing using a modifying means, a comparator means, a formation means, and an information retrieval means.

U.S. Patent No. 3,568,156 to Thompson, "Text Matching Algorithm"

Thompson is understood to disclose a method and system for searching for matching strings of elemental units in a parallel fashion. Each character in the string is assumed to be the beginning of a possible matching expression and is tested for that possibility. If a match occurs, an identification of the next character to be matched is stored. If a match has already occurred, the character is also matched against the next possible characters in the sub-string. The sub-string disclosed in Thompson refers to the remaining characters in the string, when the character being compared is not the first character in the string, since each character is assumed to be the beginning of a possible matching expression. When the tests for all possible matches are completed, the next character is tested and the stored identifications tested one at a time.

However, Thompson does not teach the invention of claims 1, 12, 23, or 34 and in particular does not teach any technique for forming a plurality of sub-strings or using an information retrieval technique on the sub-strings to determine a known string of characters. Even though Thompson discloses a system and method for finding an exact match between two strings of characters, Thompson does not disclose a system and method for text equivalencing. The text equivalencing engine of the present invention is capable of determining when one string of characters or text is equivalent or probably equivalent to another string of characters or text. The present invention text equivalencing engine can determine not only when an exact match occurs, but also when two pieces of text are equivalent.

**PATENT**

Thompson's text matching algorithm provides an organizational technique for testing character strings in parallel. Thompson does not disclose, responsive to not finding a match, forming a plurality of sub-strings of characters. Furthermore, Thompson fails to disclose using an information retrieval technique on the sub-strings to determine equivalent text.

U.S. Patent No. 6,131,082 to Hargrave et al. "Machine Assisted Translation Tools Utilizing an Inverted Index and List of Letter N-Grams"

Hargrave et al. is understood to disclose a method and system for machine assisted translation for translating from a source language to a target language. The system and method uses a translation memory including four main modules: a language analyzer module, an entropy weight normalization module, an indexing module, and a retriever module. The language analyzer module determines unique letter n-grams and weights each n-gram denoting the n-gram's importance. The entropy weight normalization module normalizes the weights to remove the bias of the longer segments. The indexing module creates a table of n-grams with a list of the text segments that contain them. The retriever module uses the translation memory to quickly find text segments similar to a query text segment.

However, Hargrave et al. does not disclose the invention of claims 1, 12, 23, or 34 and in particular does not teach modifying a string of characters using a predetermined set of heuristics. Even though Hargrave et al. discloses a language analyzer module for determining unique letter n-grams, Hargrave et al. does not disclose modifying the string of characters. The language analyzer module disclosed in Hargrave et al. merely forms word fragments or n-grams. However, in Hargrave et al. the words are not modified using a set of heuristics as claimed herein. A compilation of the n-grams in Hargrave et al. would be identical to the

**PATENT**

original word. The present invention text equivalencing engine modifies the string of characters using a predetermined set of heuristics. In the present invention the modified string is also compared with a known string of characters. Likewise, Hargrave et al. does not teach comparing the modified string with a known string of characters nor does Hargrave et al. teach responsive to not finding a match, forming a plurality of sub-strings of characters from the string of characters. Since Hargrave et al. does not modify the string of characters, Hargrave et al. also does not compare the modified string. Rather, Hargrave et al. uses the language analyzer module to determine unique letter n-grams.

Furthermore, Hargrave et al. solves a different problem from the problem solved by the present invention. Hargrave et al. solves a problem of language independence and fuzzy retrieval in a translation memory system. In the translation memory system disclosed in Hargrave et al. the source language and target language are associated with each other by a human setting up the machine translator. The present invention finds equivalent text to one known string text, for example a song or book title. The text equivalencing engine of the present invention automatically associates an equivalent string text to the known string of text.

U.S. Patent No. 5,392,212 to Geist, "Apparatus for Identifying Unknown Words by Comparison to Known Words"

Geist is understood to disclose an apparatus for character string analysis using a first memory area for data corresponding to an unknown word, a plurality of second memory areas for storing data respectively corresponding to one of a plurality of dictionary words, a control mechanism for dividing the unknown and the known words into word fragments, an arithmetic logic unit for comparing the unknown word fragments and the known word fragments and for



**PATENT**

generating an output signal indicating when a hit occurs, and a hit counter for counting the number of hits.

However, Geist does not teach the invention of claims 1, 12, 23, or 34 and in particular does not teach any technique for using an information retrieval technique on a sub-string to determine a known string of characters. Even though Geist teaches comparing the word fragments, Geist does not teach comparing the word fragments using an information retrieval technique. The text equivalencing system of the present invention uses an information retrieval technique to determine when one sub-string is equivalent to another sub-string.

Furthermore, Geist fails to disclose modifying a string of characters using a predetermined set of heuristics. Geist discloses dividing dictionary words and unknown words into fragments and comparing the dictionary fragments to the unknown fragments. In Geist the words are simply divided into fragments, but are not modified. A compilation of the word fragments would be identical to the original word. Thus, Geist does not teach modifying the unknown fragments or the unknown words using a predetermined set of heuristics prior to comparing them to the dictionary words or dictionary word fragments.

U.S. Patent No. 5,303,150 to Kameda, "Wild-Card Word Replacement Using a Word Dictionary"

Kameda is understood to disclose a word input system that uses a word dictionary for object texts written in an alphabet and a search means to search the word dictionary, which uses a replacement means to replace a word input. The replacement means can include replacing a wild card character for one portion of a character string of the word searched by

**PATENT**

search means. Kameda discloses searching for a wild card character and when a wild card character is present, searching for a character string that includes the wild card character in the word dictionary. If the search is successful, the word is completed by replacing the wild card character.

However, Kameda does not teach the invention of claims 1, 12, 23, or 34 and in particular does not teach any technique for forming a plurality of sub-strings of characters. Kameda discloses a system for replacing a wild card character. While Kameda discloses comparing a string of characters, with a wild card character replaced, to a dictionary word, Kameda does not disclose the technique of separating an unknown word and a known word into a plurality of sub-strings. The text equivalencing engine of the present invention separates known and unknown words into sub-strings. Furthermore, Kameda fails to teach using an information retrieval technique on the sub-strings of characters to determine a known string of characters equivalent to the string of characters. Kameda merely replaces a wild card character and compares the newly formed word to a dictionary word.

U.S. Patent No. 5,241,674 to Kuorsawa et al., "Electronic Dictionary System with Automatic Extraction and Recognition of Letter Pattern Series to Speed up the Dictionary Lookup Operation"

Kuorsawa et al. is understood to disclose a system and method for an electronic dictionary using an input means for entering a document image, a recognition means for automatically extracting and recognizing each letter pattern entered, display means for displaying the entered input document with an enclosure image enclosing each letter pattern series for the extracted letter patterns, a user operation means for allowing a user to specify a desire done of the letter pattern enclosed, and a look up means for looking up information on a

**PATENT**

letter code series corresponding to the letter pattern series. The look up means is a search of a data base containing prepared information on a number of letter patterns and letter codes.

However, Kuorsawa et al. does not teach the invention of claims 1, 12, 23, or 34 and in particular not teach any technique for text equivalencing. The text equivalencing engine of the present invention is capable of determining when one string of characters or text is equivalent or probably equivalent to another string of characters or text. The present invention text equivalencing engine can determine not only when an exact match occurs, but also when two pieces of text are equivalent. Kuorsawa discloses a database search looking for an exact match, but Kuorsawa does not disclose identifying an equivalent match. If an exact match is not present, Kuorsawa cannot provide an equivalent match or a probability of an equivalent match.

Furthermore, Kuorsawa et al. fails to disclose a technique for forming a plurality of sub-strings of characters. Kuorsawa also fails to disclose using an information retrieval technique on the sub-strings of characters. Kuorsawa discloses looking up information that corresponds to an exact match of a letter pattern. Kuorsawa compares a known letter pattern to an unknown letter pattern to determine this match. However, Kuorsawa does not separate the letter pattern into a plurality of sub-strings and does not apply an information retrieval technique to the plurality of sub-strings.

**Other Claims**


**PATENT**

Since the cited references do not disclose the claimed features discussed above, claims 1, 12, 23, and 34 are patentably distinguishable over the references. The remaining claims in the present application each incorporate the limitations of claims 1, 12, 23, or 34, and include additional features and limitations. Therefore, all claims herein are patentably distinguishable over the cited references for at least the reasons discussed above.

Consideration of the cited references and other information and grant of special status by this petition is solicited.

Respectfully submitted,

Dated: May 9, 2002

By:   
Elaine M. Heal, Reg. No.: 44,149  
Fenwick & West LLP  
Two Palo Alto Square  
Palo Alto, CA 94306  
Tel.: (650) 858-7113  
Fax.: (650) 494-1417